10多处理机系统

2019年2月13日 14:37

-
- 多处理机系统的基本概念

1. 提高计算机系统性能的主要途径：
    1) 提高元器件的运行速度，尤其是处理器芯片
    2) 改进系统的体系结构，尤其是引入多处理器
2. MPS多处理机系统Multiprocessor System是20世纪70年代出现的

一. 多处理机系统的引入

1. CPU时钟频率问题
    1) 早期提高CPU时钟频率可提供计算速度，随芯片制造工艺水平提供，CPU的时钟频率已从每秒数十次发展到数兆赫兹GHz
    2) 然而这种方法受限于CPU指令和数据以电子信号的方式通过介质送入送出的传输时间。电子信号早真空的速度为30cm/ns，在铜线或光纤中的传输速度约20cm/ns，100GHz计算机的信号路径便不得超过2mm
    3) 而元器件缩小又使散热成为一个棘手的问题，高端Pentium系统的CPU散热器体积已超过了其本身的体积
2. 增加系统吞吐量：增加处理机数目使系统在单位时间内能完成更多工作
3. 节省投资：达到相同处理能力的情况下，多处理机系统更省外设、内存等
4. 提高系统可靠性：将故障处理机的任务迁移到其他处理机的系统重构功能

二. 多处理机系统的类型

1. 紧密耦合MPS和松散耦合MPS
    1) 紧密耦合Tightly Coupled：通过高速总线或高速交叉开关实现多处理机互连，所有资源和进程都由操作系统实施控制管理
        (1) 多处理器共享整个主存和IO设备，访问时间约10~50ns
        (2) 多处理器与多存储器分别相连，或将主存划分成独立访问的模块
            i. 实现了处理机同时访问主存
            ii. 处理机间的访问采用消息通信方式，在10~50μs内发出
            iii. 互连较慢，软件实现较复杂，但使用构件方便
    2) 松散耦合Loosely Coupled：通过通道或通信线路实现互连，每台计算机有各自的存储器和IO设备，配置了OS管理本地资源和本地进程，独立工作，必要时才交换信息协调工作。消息传递一般需要10~50ms
2. 对称多处理器系统SMPS和非对称多处理器系统ASMPS
    1) Symmetric Multiprocessor System：各处理器单元的功能和结构都相同，是最常见的MPS，如IBM的SR/600 Model F50用了4片Power PC
    2) Asymmetric Multiprocessor System：有多重类型的处理单元，功能结构各不相同。系统中只有一个主处理器，有多个从处理器

-
- 多处理机系统的结构

1. 共享存储器的MPS中，若干处理器共享一个RAM，系统需要为每个CPU的程序提供一个完整的虚拟地址空间视图，每个存储器地址单元均可被所有CPU读写
    1) 方便了处理机间的通信
    2) 需在进程同步、资源管理及调度上，做出有别于单处理机系统的特殊处理
    3) 进程对不同存储器模块的读写速度存在的差异，形成了不同的结构：
        (1) Uniform Memory Access：统一/一致性内存访问
        (2) Nonuniform Memory Access：非统一/非一致性内存访问

一. UMA多处理机系统的结构

1) 各CPU的功能结构都相同，无主从之分，对各存储器单元的读写速度相同
2) 是一种SMPS，按处理机与存储器模块的连接方式不同，可分为：

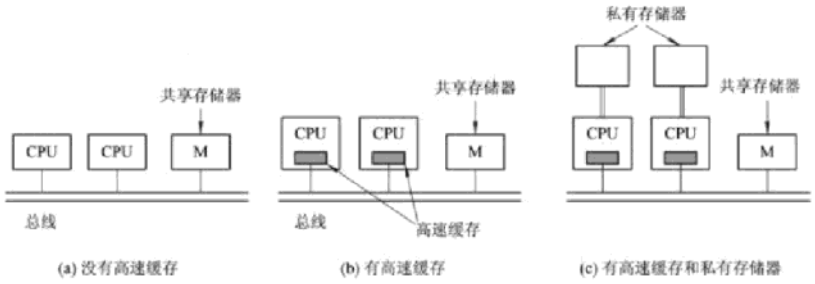
1. 基于单总线的SMP结构/均匀存储器系统
    1) 所有处理器通过公用总线访问同一物理存储器，只需运行操作系统的一个拷贝，单处理器系统的程序可直接移植，只要总线空闲，CPU即可将存储器地址放到总线上并插入若干控制信号，等待存储器将所需内容放上总线
    2) 缺点是可伸缩性有限，总线资源的瓶颈效应限制了CPU数目难以超过20
    3) 可通过在CPU内部、板上、附近等地设置高速缓存，减少其对总线的访问频率，大幅减少总线上的数据流量，支持更多CPU，高速缓存交换和存储单位一般为32或64字节块
    4) 

图10-1 基于总线的SMP结构

2. 使用多层总线的SMP结构
    1) 总线瓶颈问题的进阶解决方法是给每个CPU再配一个本地私有存储器
    2) 各CPU的本地总线负责连接私有存储器、IO设备、系统总线
    3) 系统总线一般在通信主板实现，用于访问共享存储器和连接CPU本地总线
    4) 一般只将共享变量放在共享存储器中，减少系统总线的流量，支持16~32个CPU，但这样提高了编译器的要求，编程时对数据安排的难度也高了
3. 使用单级交叉开关的系统结构
    1) 交叉开关crossbar switch类似电话交换系统，将所有CPU结点和存储器结点通过交叉开关阵列相互连接，每个交叉开关均为两个结点提供一条专用连接通道，避免了链路的争夺，方便了CPU间的通信
    2) 闭合状态称为开；打开状态称为关

3)

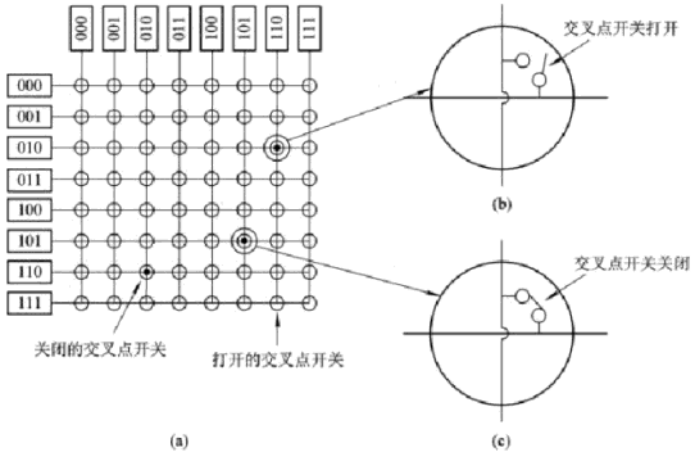


图10-2 使用交叉开关的UMA多处理机系统

4) 使用交叉开关的UMA多处理机系统的特征

(1) 结点间的连接：一般是NxN的阵列，每列都只能开一个开关

(2) CPU结点与存储器之间的连接：为支持并行存储访问，一行可同时接通多个开关

(3) 交叉开关的成本为N方，端口数N一般不超过16

4. 使用多级交换网络的系统结构

1)

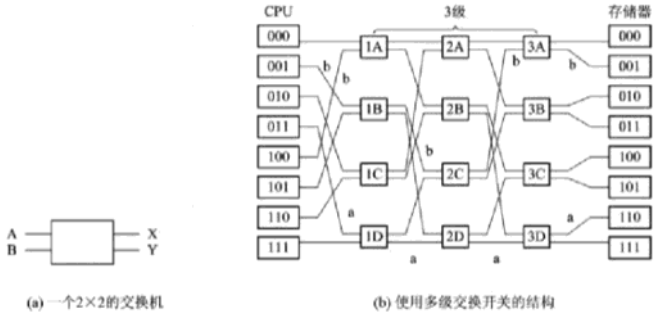


图10-3 使用多级交换网络的SMP结构示意图

2) 连接多个图a所示交叉开关即可实现多级交叉开关网络，每个最小交叉开关都是一个交叉开关级，处理机和存储器模块分别位于网络两侧

3) 所有处理机访问方式都一样，机会均等，多条路径减少了阻塞概率、分散了流量，提高了访问速度；但硬件结构昂贵，处理机数一般也不超过100

二. NUMA多处理机系统结构

1) 上述UMA的SMP体系都有共享性，随CPU数量增加，路径流量急剧增加，形成超载，使内存访问冲突迅速增加，制约了系统性能，浪费了CPU资源，降低了CPU性能的有效性，为有效扩展，需要利用NUMA技术

1. NUMA结构和特点

1) NUMA系统访问时间随存储字的位置不同而变化，公共存储器和分布在所有处理机的本地存储器共同构成了系统的全局地址空间

2) NUMA拥有多个处理机模块/节点，各节点间通过一条公用总线或互连模块进行连接和信息交互，可包含多达256个CPU

3) 各节点又可由多处理机组成，如四个拥有各自独立本地存储器、IO槽口的奔腾处理机，可通过一条局部总线和一个单独主板的群内共享存储器连接

4) NUMA结构特点：共享存储器在物理上分布式，在逻辑上连续，存储器

的集合就是全局地址空间，每个CPU都可访问所有内存，但指令不同

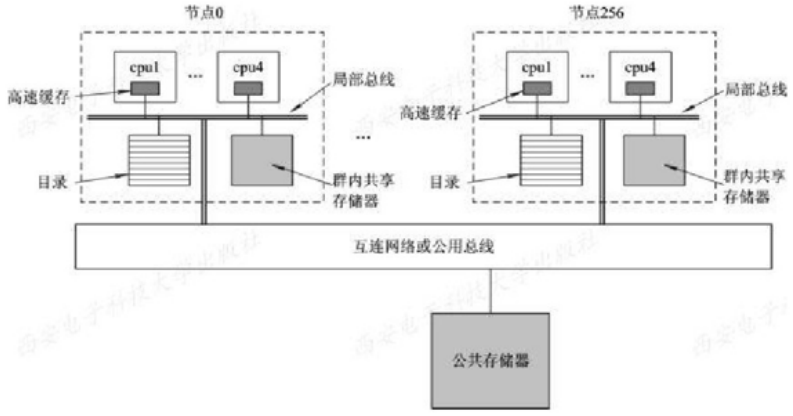
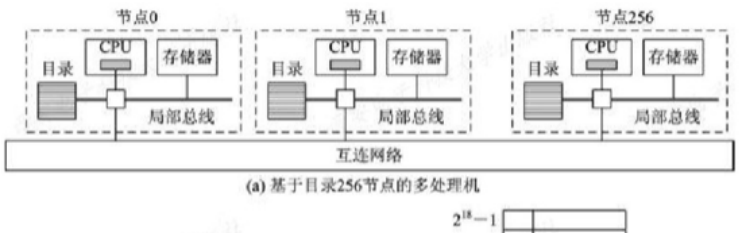
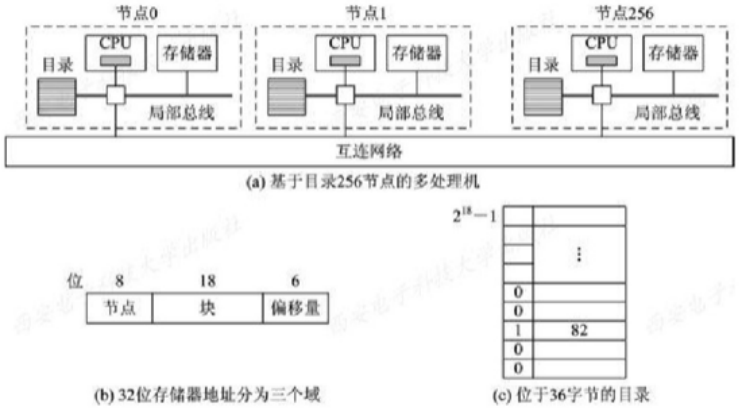
5) NUMA存储器分三层：本地存储器、群内共享存储器、全局共享存储器或其他节点存储器。访问本地存储器最快，访问属于其他处理机的远程存储器较慢。所有机器有同等访问公共全局存储器的权利，不过访问群内存储器快于访问公共存储器
6) 运行在NUMA系统的程序按址寻址数据时，首先察看本地存储器，再是本节点的共享存储器，最后是其他节点的远程内存（全局共享存储器）
7) 为更好地发挥系统性能，应尽量减少不同节点间的信息交互。各CPU都配备了高速缓存的NUMA称为CC-NUMA，没有配的称为NC-NUMA
8) 

图10-4 NUMA结构的多处理机系统

2. CC-NUMA构造方法
   1) 基于目录的多处理机思想：同一CPU拥有的若干高速缓存单元，以一定数量为一组，构成一个高速缓存块，为每个CPU配置一张高速缓存块目录表，记录和维护每个缓存块的位置和状态，读写高速缓存、变换高速缓存块节点、修改目录、访问存储器单元等操作前需要先查表
   2) 一般每个表项记录一个本地高速缓存块地址，每个高速缓存单元的内容是某个本地存储器单元内容的拷贝
   3) 32位机适合16位表项长度，将存储器空间划分为若干长度为2^(22-16)的存储器单元组，对应地目录项也应有2^(32-8-6)个
   4) 当CPU发出远程存储器单元访问指令后，由操作系统的MMU翻译出物理地址，将请求消息通过互联网络发送给对应结点，询问其对应块的地址，该节点收到请求后，将对应块内容送进请求CPU的高速缓存中
   5) 若该内容已进入其他节点的高速缓存，则会通知该节点将该内容送给发出请求的节点，之后这三个节点需要修改本地目录的对应内容指向新位置
   6) 可见，这种NUMA结果需要大量消息传递实现存储器共享，访问远程内存的延时远超本地内存，一个存储块只能进入一个节点的高速缓存限制了存储器访问速度的提高，CPU数量增加无法使系统性能线性增加



(a) 基于目录256节点的多处理机



7)

图10-5 CC-NUMA构造方法

-
- 多处理机操作系统的特征与分类

一. 多处理机操作系统的特征

1. 并行性：依赖于系统结构的改进，多处理机可使多个进程并行执行，不过每个实处理机依旧可用多道程序技术虚拟为多个虚处理机，实现并发
2. 分布性：对任何结构的多处理机系统，任务、资源及对它们的控制等方面都呈现出一点的分布性，尤其在松散耦合系统中
   1) 任务：只要能把作业分为多个可并行执行的子任务，便可分给多个处理机
   2) 资源：各处理机都可拥有存储器、IO设备等本地资源
   3) 控制：各处理单元可配置自己的操作系统，控制管理本地及协调通信共享
3. 机间的通信和同步性：不同处理机之间的同步和通信也对提高并行性、改善系统性能至关重要，且实现机制复杂得多
4. 可重构性：某处理机或存储模块等资源发生故障时，系统需要自动切除故障，换上备份资源，对系统重构，保证继续工作

二. 多处理机操作系统的功能

1. 进程管理
   1) 进程同步：需要额外解决不同处理机并行执行时引发的同步问题
   2) 进程通信：处理机间的通信信道较长，甚至要通过网络。一般是间接通信
   3) 进程调度：了解各处理机适合什么任务，各作业间的关系，哪些任务需顺序执行，哪些可并行，各任务需要什么资源，实现负载的平衡
2. 存储器管理：除了地址变换机构和虚拟存储器外，还需
   1) 地址变换机构需要确认物理地址是否远地存储器
   2) 访问冲突仲裁机构：多处理机竞争同一存储器模块时需要按规则决定顺序
   3) 数据一致性机制：确保共享主存的数据在多个本地存储器中拷贝一致
3. 文件管理
   1) 集中式：所有处理机的所有用户文件集中在某处理机的文件系统中
   2) 分散式：各处理机配置各自的文件系统，难以共享
   3) 分布式：系统中所有文件都可分布在不同处理机上，但逻辑上形成整体，用户无需了解具体物理位置即可存取。存取速度和文件保护有重要意义
4. 系统重构：自动切除故障资源并换上备份资源，若没有备份资源则重构系统是

指降级运行。若故障处理机上亟待运行的进程应安全转移至其他处理机

三. 多处理机操作系统的类型

1. 主从式master-slave
    1) 运行操作系统的特定处理机称为主处理机master processor
    2) 主处理机负责保持和记录所有处理机的属性、状态等信息，将从处理机视作可调度和分配的资源，为它们分配任务。从处理机无调度功能
    3) 工作流程：从处理机提交任务申请，等待主处理机发回应答；主处理机收到请求后中断当前任务，识别请求并转入对应处理程序，分配合适的任务给发出请求的从处理机。如CDC Cyber-170的外围处理机Po就是这种主处理机，又如DEC System10只有一台主处理机和一台从处理机
    4) 优缺点：
        (1) 易实现：只需适当扩充传统单机多道程序系统；除一些公用例程外，不需改写成可重入的，表格控制、冲突和封锁问题等都能简化
        (2) 资源利用率低：执行大量短任务时，请求任务队列较长，形成瓶颈
        (3) 安全性较差：主处理机发生不可恢复性错误易造成整个系统的崩溃
    5) 因此它只适合工作负载不重、处理机数量不多、只有一个处理机性能极高的非对称多处理机系统
2. 独立监督式separate supervisor system/独立管理程序系统
    1) 各处理机都有自己的管理程序/OS内核、IO设备和文件系统等专用资源、类似单机操作系统的管理和分配等功能的操作系统。如IBM370/158
    2) 优缺点：
        (1) 自主性强：各处理机可根据自身及任务的需要执行各种功能
        (2) 可靠性高：处理机相对独立，局部故障不会引起整个系统崩溃。不过缺乏统一管理和调度机制会使补救工作较困难
        (3) 实现复杂：管理程序的代码必须是可重入的，或者需要为每个处理机提供专用的管理程序副本；访问公用表格易发生冲突，需要仲裁
        (4) 存储空间开销大：每台处理机都有局部存储器驻留操作系统内核
        (5) 处理机负载不平衡：无主处理机负责管理和调度，难以平衡负载
3. 浮动监督式floating supervisor control mode/浮动管理程序控制方式
    1) 所有处理机组成一个处理机池，共享所有资源，某段时间内可以指定任一台（或更多台）处理机作为主处理机（组），根据需要，可以切换到其他处理机（组）。如IBM3081的MVS，C·mmp的Hydra等
    2) 优缺点：
        (1) 灵活：处理机可访问任何资源，大多数任务可在任一处理机运行
        (2) 可靠：任一处理机失效，只是少了一台可分配处理机而已，可切换
        (3) 负载均衡：可将IO中断等非专门操作分配给空闲处理机
        (4) 实现复杂：存储器模块和系统表格访问冲突需仲裁，如访问存储器用硬件解决，系统表格用优先级策略解决；多个主处理机同时执行的管理服务子程序需要有可重入性
            i.

ii.
iii.
iv.
v.
vi. ----------我是底线-------